Nucleotide Variation and Divergence in the Histone Multigene Family in *Drosophila melanogaster*

Yoshinori Matsuo[1] and Tsuneyuki Yamazaki[2]

Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

Manuscript received October 23, 1987
Accepted for publication January 24, 1989

ABSTRACT

Nucleotide differences in the histone H3 gene family in *Drosophila melanogaster* were studied on three levels: (1) within a chromosome, (2) within a population and (3) between species (*D. melanogaster* and *Drosophila simulans*). The average difference within the H3 gene within a chromosome was 0.0040 per nucleotide site, about 52% of that within a population (0.0077). The proportion of divergent sites between the two species was 0.0575, which is about 8.5 times the difference within a species. The distribution of divergence between species was similar to that of variation within a species. Divergence and variation were noted to be greatest in the 3′ noncoding region and least in the coding region. Values intermediate between these were found for the 5′ noncoding region. Divergence and variation in silent sites exceeded those in the total coding region, thus indicating possible purifying selection for amino-acid-altering change. Phylogenetic relations among H3 genes and genetic differences on these three levels are evidence for the concerted evolution of the histone gene family. The molecular mechanism by which variation is produced and maintained is discussed.

THE histone repeating unit of *Drosophila melanogaster* consists of 5 histone genes, H1, H3, H4, H2A and H2B and occurs in tandem about 110 times in the 39 D–E region of chromosome *IIR* (PARDUE *et al.* 1977; KARP 1980). Most repeating units can be classified as one of two types on the basis of length, 4.8 kb (an S unit) and 5.0 kb (an L unit) (LIFTON *et al.* 1977). The latter unit has a tRNA derived insertion sequence 240 bp in length in the spacer between H1 and H3 (MATSUO and YAMAZAKI 1989). These types can be distinguished by the presence or absence of a restriction enzyme site (*Eco*RI) in a repeating unit as well as difference in the length of such a unit. These different types are not organized randomly but are clustered (SAIGO, MILLSTEIN and THOMAS 1981; MATSUO and YAMAZAKI 1989). Some repeats are interrupted by fragments not homologous to the histone repeating unit (SAIGO, MILLSTEIN and THOMAS 1981; MATSUO and YAMAZAKI 1989). One such example is a repeating unit containing retrotransposon 297 (IKENAGA and SAIGO 1982). The molecular structure and DNA sequences of the histone gene repeating unit have been characterized by KARP (1980), GOLDBERG (1979) and MATSUO and YAMAZAKI (1989).

Population genetical studies on the DNA level are providing new information on naturally occurring variation in silent sites, introns and intergenic spacers (LANGLEY, MONTGOMERY and QUATTLEBAUM 1982; KREITMAN 1983; KREITMAN and AGUADE 1986; AQUADRO *et al.* 1986). It is now possible to get information about genetic changes occurring in speciation and evolution by comparing homologous genes from different species (BODMER and ASHBURNER 1984; COHN, THOMPSON and MOORE 1984; KREITMAN and AGUADE 1986). The single-copy *Adh* locus of *D. melanogaster* has been studied most extensively so far, but more information on other loci, duplicated genes, multigenes, transposons and repetitive sequences should be obtained for a better understanding of variation and evolution of genomes.

Members of a multigene family appear to have evolved not independently but as units such as 28S and 18S ribosomal DNA, 5S rDNA, globin genes, immunoglobulin genes, heatshock genes and histone genes. Such a process is called concerted or coincidental evolution (ARNHEIM 1983; DOVER 1982). Several molecular mechanisms for this have been proposed. Studies on the variation and divergence of a multigene family should facilitate clarification of the actual mechanism(s) involved (and the rate of this evolution).

Little is understood regarding nucleotide differences and divergence in a multigene family within a species. In this study, histone gene repeating units of *D. melanogaster* were examined on three different levels (within a chromosome between loci, within a population between chromosomes and between closely related species, *D. melanogaster* and *Drosophila simulans*). Nucleotide differences in the histone genes were determined so as to provide some clarification of the evolutionary process. The mechanism by which

[1] Present address: Department of Ecology, Ethology and Evolution, University of Illinois, Shelford Vivarium, 606 East Healey Street, Champaign, Illinois 61820.
[2] To whom correspondence should be addressed.

variation and divergence in a multigene family are maintained is discussed.

MATERIALS AND METHODS

Strains and population: The Akayu isogenic strain of *D. melanogaster* (AK194), whose two major chromosomes (*II* and *III*) were extracted simultaneously by a balanced lethal system (*Cy/Pm SbSer/Pr*, see YAMAZAKI *et al.* 1984) was used for the study of nucleotide differences within a chromosome. The Akayu cage population used to study differences within a population was established using 1600 genomes in 1974 (YAMAZAKI *et al.* 1983). At least 3000 flies have consistently been maintained in the cage population for the past 11 yr. The population is polymorphic at various enzyme loci, but monomorphic with respect to universal inversions. A *D. simulans* strain of unknown origin was obtained from the National Institute of Genetics.

Extraction of high molecular weight DNA: About 1 g of adult flies was frozen with liquid N_2, homogenized and suspended in 4 ml of buffer [50 mM Tris-HCl (pH 8.0)/10 mM EDTA (pH 8.0)/10 mM NaCl] with 40 μl of Proteinase-K solution (20 mg/ml). After incubation for 4 to 6 hr at 50°, 3 g of CsCl (per 4 ml solution) and 250 μl of EtBr (10 mg/ml) were added and mixed with the solution, followed by centrifugation at 38,000 rpm (approximately 100,000 × *g*) for 48 hr. EtBr in the collected DNA solution was removed by washing several times with isopropanol saturated with 5 M NaCl/10 mM Tris-HCl (pH 8.0)/1 mM EDTA (pH 8.0). The DNA solution was dialyzed with several thousand times its volume of TE solution (10 mM Tris-HCl/1 mM EDTA). The DNA was precipitated by ethanol, washed with 70% ethanol, and suspended in 0.1 × TE solution.

DNA preparation from plasmids and phages: DNA of the plasmids and phages was prepared by Alkali-SDS and liquid culture, respectively (MANIATIS, FRITSCH and SAMBROOK 1982). The method for DNA preparation from plasmids was as follows. An aliquot of 1.5 ml of an overnight culture was centrifuged and the collected bacteria suspended in 60 μl of buffer [50 mM glucose/25 mM Tris-HCl (pH 8.0)/10 mM EDTA]. An aliquot of 40 μl of freshly prepared lysozyme solution [50 mM glucose/25 mM Tris-HCl (pH 8.0)/10 mM EDTA/10 mg/ml lysozyme] was mixed with the bacterial solution which was then maintained for 5 min at room temperature. This was followed by the addition of 200 μl of 0.2 N NaOH/1% SDS and the solution was kept on ice for 5 min. After being mixed with 150 μl of 5 M potassium acetate (pH 4.8), the solution was stored on ice for at least 5 min. After 10 min of centrifugation, the supernatant was treated with phenol. Plasmid DNA was precipitated with ethanol, washed with 70% ethanol and suspended in 50 μl of TE solution (pH 8.0).

The liquid culture method for DNA preparation from phages was conducted as follows. About 1 μl of phage stock solution (about 10^8 phage/ml) was mixed with 0.5 ml of an overnight culture of bacteria (m.o.i., about 0.01). After incubation for 20 min at 30°, 10 ml of T broth containing 10 mM $MgCl_2$ were added to the bacterial culture followed by shaking for about 6 hr at 37°. After becoming clear, the culture was treated with 0.1 ml of chloroform followed by incubation for an additional 15 min at 37°. After centrifugation at 7000 rpm for 10 min, the supernatant was treated with RNase A (10 μg/ml) and DNase I (1 μg/ml) for 30 min at 37°. NaCl and PEG6000 were added and mixed with the DNA solution at final concentrations of 1 M and 10%, respectively. The solution was placed on ice for at least 90 min. The precipitated bacteriophage particles were recovered by centrifugation at 10,000 rpm for 15 min at 4° followed by the addition of 0.5 ml of SM solution [0.1 M NaCl/0.008 M $MgSO_4 \cdot 7H_2O$/0.05 M Tris-HCl (pH 7.5)/0.1% gelatin] and the phage solution was treated with phenol two times. The phage DNA was precipitated with ethanol, washed with 70% ethanol and suspended in 50 μl of 0.1 × TE solution.

DNA cloning: DNA fragments digested by restriction enzymes were ligated into the polylinker site of the pUC9 plasmid (VIEIRA and MESSING 1982). Insert DNA and vector DNA in 20 μl of ligation solution (0.06 mM ATP/66 mM Tris-HCl (pH 7.6)/6.6 mM $MgCl_2$/10 mM dithiothreitol DTT) were incubated overnight with 0.01 unit of T4 DNA ligase at 4°.

Transformation was conducted by the $CaCl_2$ method (MANDEL and HIGA 1970). An overnight culture of JM83 or JB1 was added to 100 times the volume of L broth and allowed to grow for about 2 hr at 37°. After the culture had been kept on ice for 10 min, the bacteria were washed with 10 ml of ice-cold buffer [50 mM $CaCl_2$/10 mM Tris-HCl (pH 8.0)], suspended in 1/15 the volume of the above buffer and kept on ice for 3 to 6 hr. An aliquot of 0.2 ml of the above competent cell solution was mixed with 1 μl of ligation mixture and kept on ice for 30 min. After incubation for 2 min at 42°, 1 ml of the L broth was added to the mixture followed by incubation for 1 hr at 37°. About 100 μl of reaction solution were spread on the media containing Amp (50 μg/ml) and X-gal (30 μg/ml) (MANIATIS, FRITSCH and SAMBROOK 1982).

Gene libraries: The gene library of the isogenic strain (AK-194) was constructed according to the following two methods ("A" [MANIATIS *et al.* (1978)] and "B" [KARN, BRENNER and BARNETT (1983)]). *Method A*: Drosophila DNA digested partially by *Eco*RI and vector DNA (charon 4) digested by *Eco*RI were separated by sucrose density gradient centrifugation (MANIATIS, FRITSCH and SAMBROOK 1982). An appropriate DNA fraction was collected, dialyzed in TE solution and precipitated with ethanol. The cohesive ends of the vector were annealed in TE solution containing 0.01 M $MgCl_2$ for 1 hr at 42°. About 2 μg of vector DNA and 1 μg of Drosophila DNA (10–22 kb) were ligated overnight with T4 DNA ligase at 4°. After portions of the reaction mixture were checked by agarose gel electrophoresis, DNA was packaged into phage particles *in vitro* by the method of HOHN and MURRAY (1977). *Method B:* Drosophila DNA digested partially by *Bam*HI and vector DNA (λ EMBL4) digested by *Bam*HI were ligated with T4 DNA ligase overnight at 4°. The phages packaged *in vitro* containing Drosophila DNA were deleted in red and gamma genes, so that recombinant phages could be selected on P2 lysogents (Q359). The *Bam*HI library was necessary to clone histone S units since S units do not have an *Eco*RI site. Using these two gene libraries it should be possible to observe genetic variation in the histone gene family between loci within a chromosome.

The gene library of the Akayu population was constructed by *Mbo*I partial digestion using the method of KARN, BRENNER and BARNETT (1983) as described above. Thousands of eggs were sampled from this population. Adult flies that had developed from the eggs were used for the DNA preparation. Thus, most of the variation of histone gene should originate from different chromosomes. This is because of the negligibly small probability ($1/2N$) that two randomly sampled histone gene units from this library derive from the same chromosome (N is the number of adults sampled, and was at least 1000). In any *D. melanogaster* library, any two histone-positive clones are likely to derive from different loci (or repeating units). The probability that

any two clones derive from the same locus is less than 0.01 (1/110 where 110 is the approximate copy number of repeating units in *D. melanogaster*). The gene library of *D. simulans* was constructed by *Eco*RI partial digestion using the method of MANIATIS *et al.* (1978) as described above.

Labeling of DNA: Nick translation was conducted following RIGBY *et al.* (1977). The *Eco*RI 5-kb fragment of pKSL100, a single repeating unit of the histone gene (SAIGO, MILLSTEIN and THOMAS 1981), was incubated with DNase I (2 ng/ml) in nick translation buffer [50 mM Tris-HCl (pH 7.5)/10 mM $MgSO_4$/1 mM DTT/5 μM dATP/5 μM dGTP/5 μM dTTP/25 μCi α-[^{32}P]dCTP] for 15 min at 37°. After the addition of DNA polymerase I (2–3 units), the solution was incubated for 2 hr more at 15°. Three volumes of stop solution [0.3 M Tris-HCl (pH 8.0)/10 mM EDTA] were added to the reaction mixture which was then treated with phenol or chloroform-isoamyl alcohol (24:1) solution. DNA was precipitated by ethanol with carrier tRNA, washed with 70% ethanol and suspended in TE solution.

Plaque hybridization: Plaque hybridization was performed according to BENTON and DAVIS (1977). Phage DNA was transferred to a nylon membrane filter (Pall's Biodyne), treated with denaturation buffer (0.5 N NaOH/1.5 M NaCl) for 20 sec and then with neutralization buffer [3 M Na-acetate (pH 5.5)] for 20 sec. The membrane filter was dried and baked for 2 hr at 80°. After prehybridization with 50 μg/ml of heat denatured salmon sperm DNA in 4 × SETDSP [1 × SETDSP: 1 × SET/10 × Denhart/0.1% SDS/0.1% Na-pyrophosphate; 1 × SET: 0.15 M NaCl/30 mM Tris-HCl (pH 7.5)/2 mM EDTA; 1 × Denhart: 0.02% bovine serum albumin (fraction V)/0.02% polyvinylpyrrolidone/0.02% Ficoll] for 1 hr at 68°, phage DNA was hybridized with the histone gene probe in the same solution as that for prehybridization for 24 to 36 hr at 68°. Washing conditions were as follows: 4 × SETDSP once, 3 × SETSP (3 × SET/0.1% SDS/0.1% Na-pyrophosphate) 3 times and 1 × SETSP once, all at 68°.

Southern blotting: Digested DNA was separated by agarose gel electrophoresis, denatured for 30 min and neutralized for 30 min. The DNA was then transferred to a nylon membrane filter following SOUTHERN (1975). The filter was dried and baked for 1 hr at 80°. After prehybridization with heat denatured salmon sperm DNA in 6 × SSC (20 × SSC: 3 M NaCl/0.3 M Na-citrate) for 1 hr at 68°, hybridization was conducted with the histone gene probe in the same solution as that for prehybridization for 24 to 36 hr at 68°. Washing conditions were as follows: 6 × SSC once, 4 × SSC twice and 1 × SSC once, all at 68°.

DNA sequencing: The DNA sequence was determined by the dideoxy chain termination method of SANGER, NICKLEN and COULSON (1977). Following the method of HATTORI, HIDAKA and SASAKI (1985), the denatured plasmids were used as a template. Plasmid DNA extracted from 1.5 ml of overnight culture was treated with 1 μl of RNase A (10 mg/ml) for 30 min at 37°. An aliquot of 30 μl of 20% PEG/2.5 M NaCl solution was mixed with 50 μl of DNA solution and kept on ice for 1 to 12 hr. After centrifuging at 12,000 rpm for 10 min, the precipitates were rinsed with 70% ethanol, dried, and suspended in 50 μl of TE buffer. An aliquot of 18 μl of the above DNA solution was mixed with 2 μl of 2 N NaOH and kept at room temperature for 5 min. The solution was then mixed with 8 μl of 5 M ammonium acetate (pH 7.4). DNA was precipitated by ethanol, rinsed with 70% ethanol, and dried thoroughly. The sequencing was performed essentially according to the procedure commonly used for M13 phage vectors (SANGER *et al.* 1980) using a Takara sequencing kit and Amersham's universal and reverse primers. Electrophoresis was conducted with 8% and 5% acrylamide-7 M urea gels. After electrophoresis, each gel was fixed in 10% methanol/10% acetic acid solution for 15 min, transferred to paper, dried, and subjected to autoradiography.

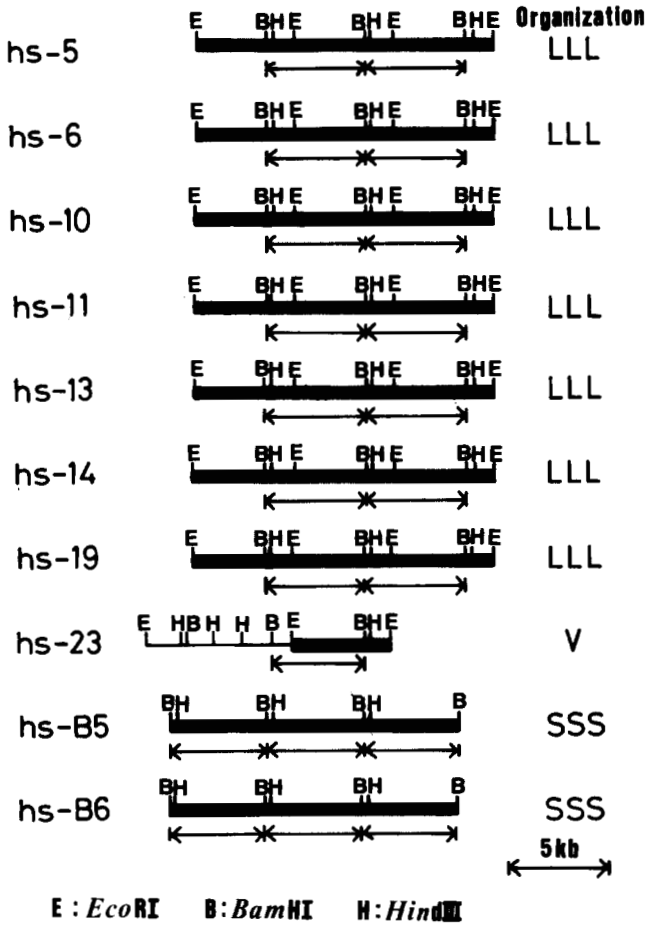

FIGURE 1.—Restriction maps of 10 clones from AK-194 isogenic strain (hs-). The organization of a histone repeating unit (L or S or variant) is shown on the right side. Possible fragments for cloning by *Bam*HI are indicated by arrows. hs-B5 and hs-B6 were obtained from the *Bam*HI library and the remaining eight clones from the *Eco*RI library.

Synthetic DNA: A 17-mer synthetic DNA (5′-GCGATGACGCTTGGCGC-3′) was prepared as the sequencing primer, whose location is shown in Figure 5.

RESULTS

Genetic variability among loci of H3 repeating units within a chromosome: Ten histone-positive clones were chosen randomly from the AK-194 library. The probability that any two repeating units from different clones derive from different repeating units is 109/110 (assuming 110 repeating units on a chromosome). The restriction maps of these clones were determined and are shown in Figure 1. Recombinant phages had one to three histone repeating units, one of which was randomly subcloned into the *Bam*HI site of plasmid pUC9. The DNA sequences of H3 coding and the 5′ and 3′ flanking regions were determined by the strategy depicted in Figure 2. The

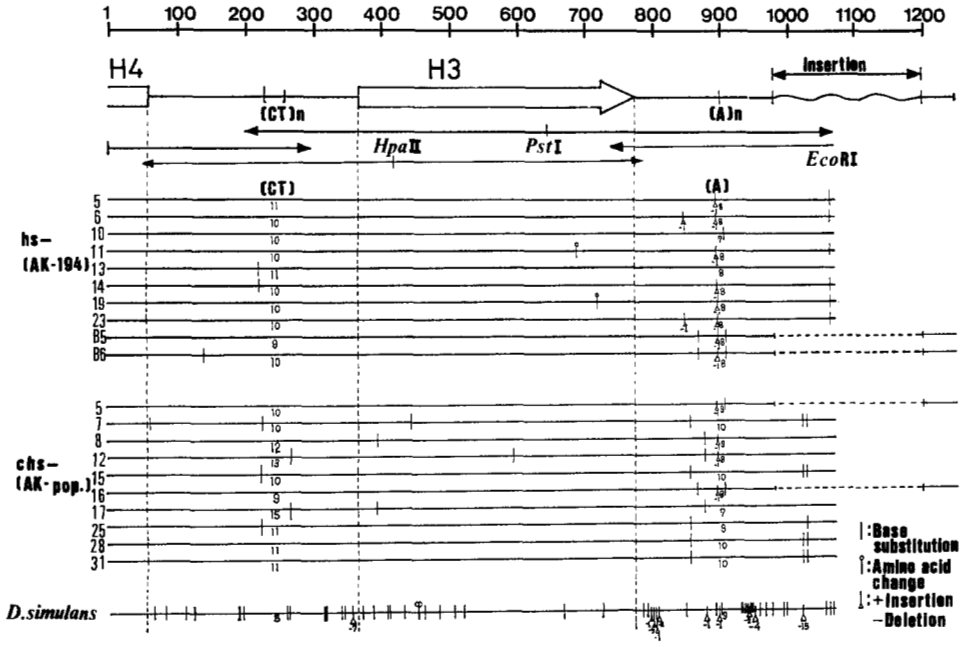

FIGURE 2.—Sequence strategy and polymorphic or divergent sites in 10 clones from a single chromosome (hs-) and 10 clones from different chromosomes (Chs-) in *D. melanogaster* and a clone from *D. simulans*. Sequencings were conducted from *Pst*I, *Hpa*II and *Eco*RI sites in addition to the synthetic primer within H4 coding region (5′-GCGATGACGCTTGGCGC-3′). Differences in the consensus sequence (Figure 5) are shown. The number of (*CT*) and (*A*) is shown in the corresponding sites along with that of deletion and insertion bases. The location of sequence corresponds to that in Figure 5.

different sites in consensus sequences and summarized data are shown in Figure 2 and Table 1.

Four differences, *i.e.*, base substitution, a few bp insertion or deletion, tRNA derived element insertion or deletion (MATSUO and YAMAZAKI 1989), and different repetitions of a simple sequence [$(CT)_n$ and $(A)_n$] were found. Two substitutions altering amino acids were observed in the clones examined. All insertions/deletions were found in the 3′ region of H3. The molecular structure and characterization of the tRNA derived element found in the spacer between H1 and H3 have already been described by MATSUO and YAMAZAKI (1989). Different repetitions of *CT* (9–11) and *A* (7–8) were found in the 5′- and 3′-regions of H3, respectively (Table 1). Although the organization of the hs-23 clone, the interrupted type, differed from the tandem cluster type (MATSUO and YAMAZAKI 1989), the number of different nucleotides in the consensus sequence was only two. Very divergent histone genes such as the late histone in the sea urchin (CHILDS *et al.* 1982) or pseudogene sequences could not be found in the 10 sampled clones, indicating the frequency of these types to be quite low or nonexistent.

The proportion of polymorphic sites and average heterozygosity of the ten clones from an isogenic strain were calculated as 0.0129 and 0.0040, respectively. In calculating these values, each of the 4 differences (see above for explanation) was counted as a single event (Table 1). The average heterozygosity used in Table 1 is the average over all nucleotide sites specified. The average heterozygosity at each nucleotide site is the average of the average heterozygosity of random combinations of 10 chromosomes at each nucleotide site, defined by $\{n/(n-1)\}(1-\sum x_i^2)$ where n indicates sample size (10 in this case) and x_i indicates sample frequency of the ith variant at each site. The average heterozygosity values of the H3 gene among the members of a family in the 5′ region, coding region, 3′ region and insertion region was 0.0039, 0.0019, 0.0072 and 0.0082, respectively. The genetic variability in the coding region was less than that in the noncoding region. There was greater variation in the 3′ than 5′ region. Similar results were obtained for the proportion of polymorphic sites.

Genetic variability in the histone gene among chromosomes of a population: Ten clones considered to have derived from different repeating units among different chromosomes (see MATERIALS AND METHODS), were chosen randomly from the Akayu cage library and their restriction maps were determined (Figure 3). In the same manner as within a chromosome, one repeating unit in a phage clone was subcloned and the DNA sequences of the H3 region were determined. The same four differences noted within a chromosome were also apparent (Figure 2 and Table 1), but no amino acid altering change was evident. The range of $(CT)_n$ and $(A)_n$ repetition was 9 to 15 and 7 to 10, respectively. These values were slightly larger than those within a chromosome.

The proportion of polymorphic sites and average heterozygosity of ten clones from a population were 0.0159 and 0.0077, respectively (Table 1). The amount of heterozygosity of the H3 gene in this study was almost the same as that of the whole histone gene repeating unit (0.009; our unpublished data) and not far from that of the Adh region (0.004–0.006; LANGLEY, MONTGOMERY and QUATTLEBAUM 1982; KREITMAN 1983). The average heterozygosity of the histone H3 gene within a population in the 5′ region, coding

TABLE 1

Nucleotide differences within a chromosome (hs-) and between chromosomes (Chs-) of H_3 and H_4 genes

	H4 coding	5′ region									H3 coding						3′ region									Insertion				
Location:		65	143	222	223	227	229	258–229	270		396	447	597	689	718		847	855	866	876	895	896–905	907			976	1021	1027	1062		
Consensus:		C	T	T	G	T	(CT)n	C			A	C	G	C	G		A	C	A	T	T	(A)n	A			In	C	A	A		
hs-5		.	.	.	.	.	11	.			.	.	.	.	.		.	.	.	.	–	8	.			.	.	.	G		
hs-6		.	.	.	.	.	10	.			.	.	.	.	.		–	.	.	.	–	8	.			.	.	.	G		
hs-10		.	.	.	.	.	10	.			.	.	.	.	.		.	.	.	.	.	7	T			.	.	.	.		
hs-11		.	.	.	.	.	10	.			.	.	.	T	.		.	.	.	.	–	8	.			.	.	.	G		
hs-13		.	.	C	.	.	11	.			.	.	.	.	.		.	.	.	.	.	8	.			.	.	.	.		
hs-14		.	.	.	T	.	10	.			.	.	.	.	.		.	.	.	.	–	8	.			.	.	.	G		
hs-19		.	.	.	.	.	10	.			.	.	.	.	A		.	.	.	.	–	8	.			.	.	.	G		
hs-23		.	.	.	.	.	10	.			.	.	.	.	.		–	.	.	.	–	8	.			.	.	.	G		
hs-B5		.	.	.	.	.	9	.			.	.	.	.	.		.	.	T	.	–	8	T			–	*	*	*		
hs-B6		.	C	.	.	.	10	.			.	.	.	.	.		.	.	T	.	–	8	T			–	*	*	*		Total
P.P.	0				0.0141							0.0049								0.0260							0.0213			0.0129
	(0/33)				(4/283)							(2/405)								(5/192)							(2/94)			(13/1007)
P.P.S.	0											0																		
	(0/9.7)											(0/103.6)																		
A.H.	0				0.0039							0.0019								0.0072							0.0082			0.0040
A.H.S.	0											0																		
Chs-5		.	.	.	.	.	10	.			.	.	.	.	.		.	.	.	.	–	9	T			–	*	*	*		
Chs-7		T	.	.	.	C	10	.			.	T	.	.	.		.	T	.	.	.	10	.			.	A	G	.		
Chs-8		.	.	.	.	.	12	.			G	.	.	.	.		.	.	.	G	–	9	.			.	.	.	.		
Chs-12		.	.	.	.	.	13	G			.	.	A	.	.		.	.	.	G	–	8	.			.	.	.	.		
Chs-15		.	.	.	.	C	10	.			.	.	.	.	.		.	T	.	.	.	10	.			.	A	G	.		
Chs-16		.	.	.	.	.	9	.			.	.	.	.	.		.	.	T	.	–	8	T			–	*	*	*		
Chs-17		.	.	.	.	.	15	G			G	.	.	.	.		.	.	.	G	.	7	.			.	.	.	.		
Chs-25		.	.	.	.	C	11	.			.	.	.	.	.		.	T	.	.	.	9	.			.	.	G	.		
Chs-28		.	.	.	.	.	11	.			.	.	.	.	.		.	T	.	.	.	10	.			.	A	G	.		
Chs-31		.	.	.	.	.	11	.			.	.	.	.	.		.	T	.	.	.	10	.			.	A	G	.		Total
P.P.	0				0.0141							0.0074								0.0313							0.0319			0.0159
	(0/33)				(4/283)							(3/405)								(6/192)							(3/94)			(16/1007)
P.P.S.	0											0.0290																		
	(0/9.7)											(3/103.6)																		
A.H.	0				0.0067							0.0028								0.0150							0.0193			0.0077
A.H.S.	0											0.0073																		
T.P.P.	0				0.0247							0.0123								0.0365							0.0319			0.0218
	(0/33)				(7/283)							(5/405)								(7/192)							(3/94)			(22/1007)
T.P.P.S.	0											0.0290																		
	(0/9.7)											(3/103.6)																		

Additions and deletions of more than one contiguous base are counted as one change. P.P., proportion of polymorphic site; P.P.S., proportion of polymorphic silent site; A.H., average heterozygosity; A.H.S., average heterozygosity of silent site; T.P.P., total proportion of polymorphic site; and T.P.P.S., total proportion of polymorphic silent site.

region, 3′ region and insertion region was 0.0067, 0.0028, 0.0150, and 0.0193, respectively. Although the degree of variation in the histone H3 gene within a population was about 1.7 times that within a chromosome, the distribution pattern of variation in the above two parameters was similar. More genetic variation was found in the noncoding region than in the coding region. Variation in the 3′ region exceeded that in the 5′ region. The average heterozygosity of silent sites in the coding region (0.0073) was higher than that of the total coding region (0.0028) and essentially the same as that of either the 3′ region (0.0150) or insertion region (0.0193). The selective constraint of silent site is found to be weak compared to that of amino acid altering sites. The 3′ and insertion regions appeared to have selectively less constraint than the coding region.

Phylogenetic relationships in the histone multigene family in *D. melanogaster*: To gain some understanding of the evolutionary relationships in the histone multigene family, two types of phylogenetic trees for 20 histone H3 genes [10 from an isogenic strain (within a chromosome, hs-) and 10 from a population (among chromosomes, chs-)], were con-

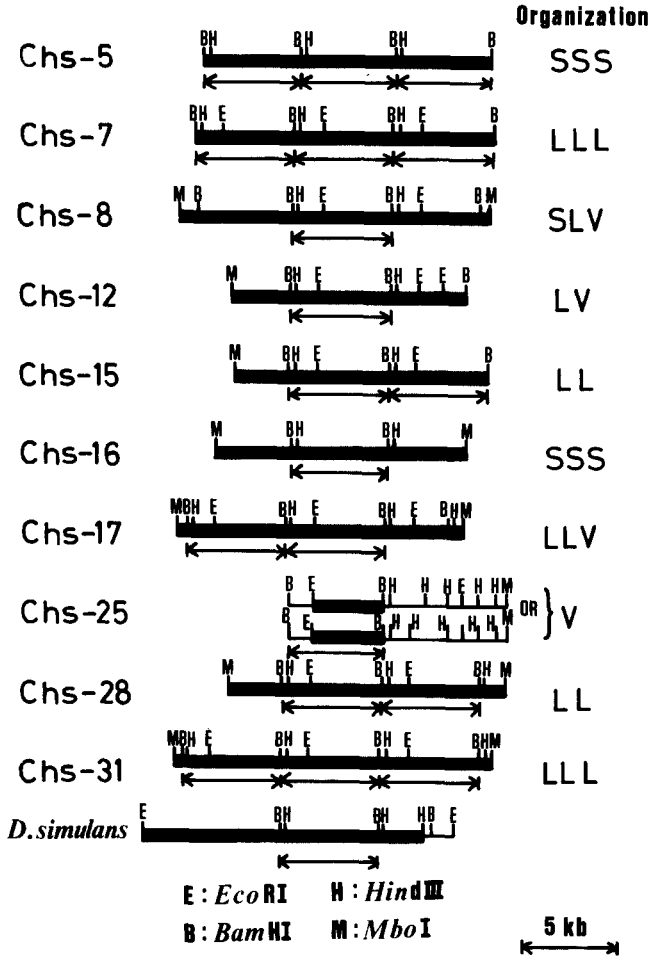

FIGURE 3.—Restriction maps of 10 clones from the Akayu cage (Chs-) and a clone from *D. simulans*.

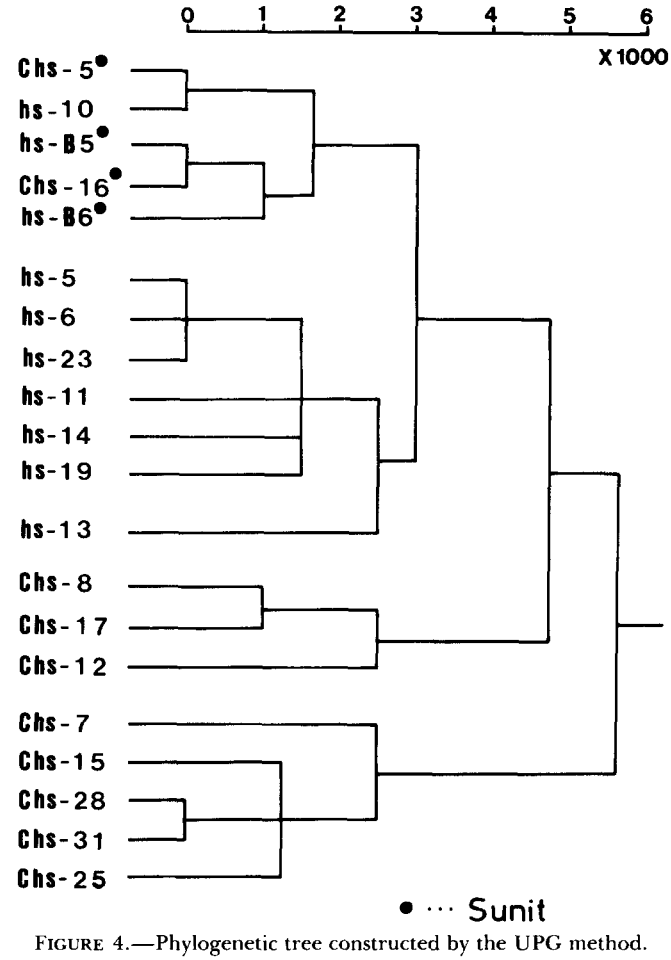

FIGURE 4.—Phylogenetic tree constructed by the UPG method. Insertions and deletions are included. The number of nucleotide substitutions per site (K') was estimated by JUKES and CANTOR's formula (1969). $K' = -(3/4)\ln[1 - (4/3)K]$, where K is the proportion of different nucleotides.

structed by the UPG method (SNEATH and SOKAL 1973): in one there is only base substitution and in the other both base substitution and insertion/deletion. Results obtained for the latter are shown in Figure 4 since the results obtained from either method were essentially the same. The locations of hs-10 and hs-13 could not be clearly determined, but the topology of (hs-5, hs-6, hs-11, hs-14, hs-19, hs-23), (chs-7, chs-15, chs-25, chs-28, chs-31) and the grouping of (hs-B5, hs-B6, chs-5, chs-16), (chs-8, chs-12, chs-17) were the same for both categories.

With regard to these phylogenetic relationship, two points in particular should be noted, First, S units showed greater similarity to each other than the S and L units or the L units with themselves. Thus, an ancestral S unit may possibly have spread to a genome and then to a population. Second, the histone genes within a chromosome showed less variation and certain mechanisms such as unequal crossing over or gene conversion must be responsible for the exchange of genetic information between histone gene clusters (between loci).

Genetic divergence of the H3 gene between *D. melanogaster* and *D. simulans*: A histone-positive clone from the *D. simulans* library was obtained and restriction maps were determined (Figure 3). The DNA sequence of the H3 region in *D. simulans* was determined in the same manner as that of *D. melanogaster* clones (Figure 5). The different sites in the two species are indicated in Figure 2. Three kinds of divergence were found in the H3 gene region between *D. melanogaster* and *D. simulans*: base substitution, short length insertion/deletion and a number of repetitions of a simple sequence (Note that the term "divergence" is used when the difference between species is discussed). The results of Southern blotting showed most of the histone repeating units in *D. simulans* to be essentially the same size (our unpublished data). Since one of the histone repeating units of *D. simulans* had a tRNA-derived insertion sequence, most of the histone repeating units in *D. simulans* should have that sequence. Still there is the possibility that the histone repeating unit without a tRNA-derived sequence exists rather infrequently. The proportion of insertion or deletion relative to total difference was higher between *D. melanogaster*

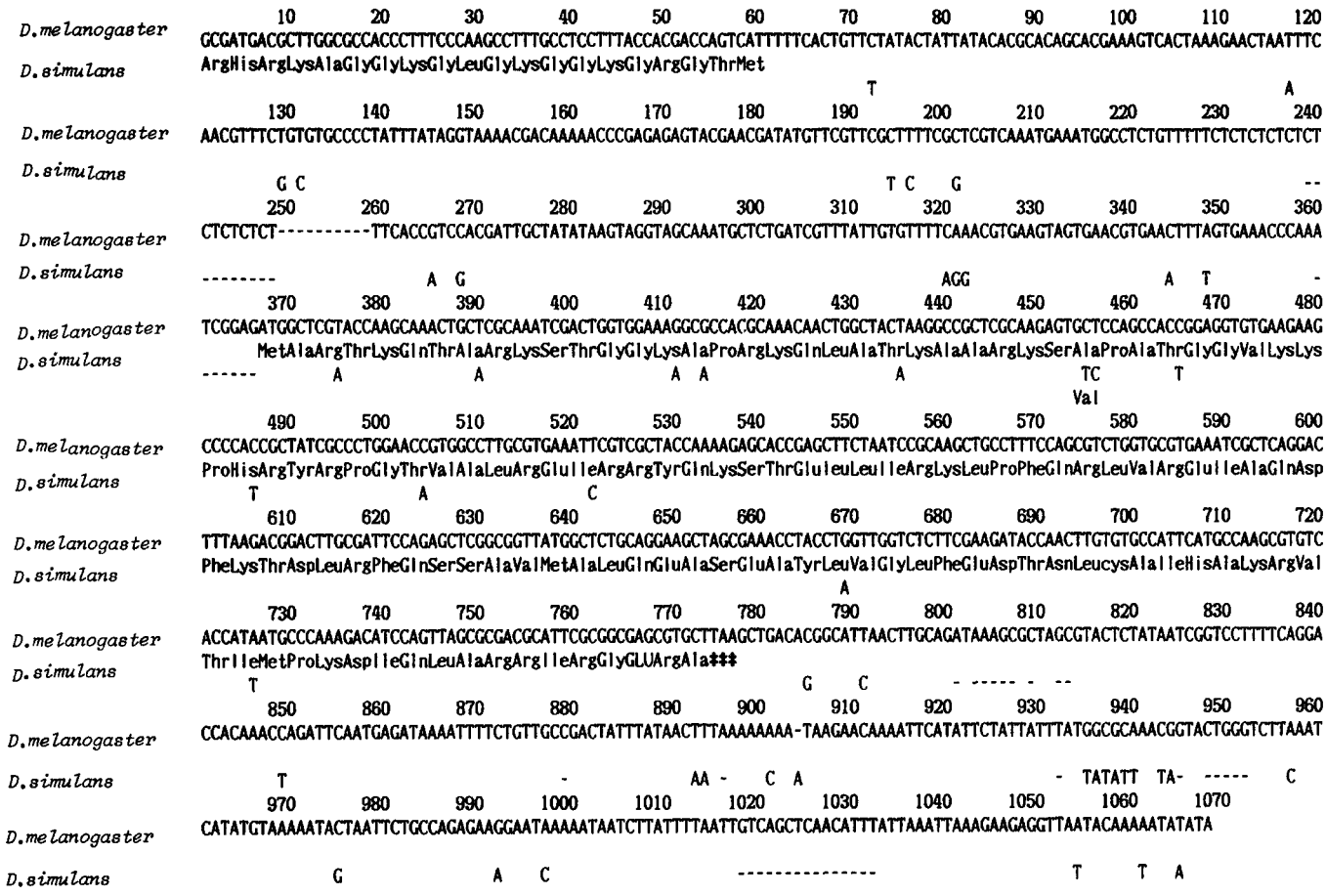

FIGURE 5.—Consensus sequence of H3 and flanking region in *D. melanogaster* and sequence from *D. simulans*. Only differences in the consensus sequence are shown. Deletion is indicated by "-". The location No. 1 is the first base of synthetic DNA (see MATERIALS AND METHODS). Sequence analysis can be done from the positions 25 to 1068 in the case of the L unit.

and *D. simulans* (0.231; 12/52) than that within the former species (0.133; 2/15). Most of the insertions or deletions occurred in the 3′ region of the H3 gene. The number of repetitions of *CT* in *D. simulans* was 5, which is quite different from that found in *D. melanogaster* (9–15).

The average proportion of divergent sites between all of *D. melanogaster* and *D. simulans* was 0.058 and 0.110 in the total and silent sites, respectively (Table 2). These values were about 8.5 and 15.1 times the differences within-species variation in the total and silent sites, respectively. The differences in the 3′ region, 5′ region and coding region were 0.100, 0.063 and 0.033, respectively. Divergence occurred more frequently in the noncoding than coding region as did also genetic variation within a species (Table 2). The less functional site or silent site evolved more rapidly than the total coding region.

Concerted evolution of the histone gene family in Drosophila: From analysis by SOUTHERN blotting, about the same copy number of histone genes appeared present in several Drosophila species (COEN, STRACHAN and DOVER 1982; our unpublished data). It thus seems reasonable to assume that histone gene amplification occurred prior to the divergence of these species. It was considered that if mutation accumulates between a pair of histone genes independently without the exchange of genetic information between them, it should be possible to find the same number of differences among histone gene copies within a chromosome, among nonallelic histone genes in different genomes and among species (DOVER 1982; ARNHEIM 1983).

The degree of difference in histone H3 genes between species and within a population was about 14.0 and 1.7 times that within a chromosome, indicating the possibility of concerted evolution of the histone gene family. Observed phylogenetic relationships also confirmed this possibility.

DISCUSSION

Genetic differences in the histone gene family in *D. melanogaster* were studied at three different levels, within a chromosome, within a population and between closely related species (*D. melanogaster* and *D. simulans*). About 52% of genetic variation within a population was found within a chromosome. Multiple copies may be efficient not only for producing many

TABLE 2

Proportion of polymorphic sites and average heterozygosity within population and between *D. melanogaster* and *D. simulans*

Property[a]	H4 coding	5′ region	H3 coding	3′ region	Insertion	Total
Proportion of polymorphic site	0 (0/33)	0.025 (7/283)	0.012 (5/405)	0.036 (7/192)	0.043 (4/94)	0.023 (23/1007)
Proportion of polymorphic silent site	0 (0/9.7)		0.029 (3/103.6)			
Average heterozygosity	0	0.007	0.002	0.015	0.012	0.007
Average heterozygosity of silent site	0		0.008			
Proportion of divergent site between *D. melanogaster* and *D. simulans*	0	0.063	0.033	0.100	0.100	0.058
Proportion of divergent silent site between *D. melanogaster* and *D. simulans*	0		0.110			

[a] "Polymorphic" and "heterozygosity" were used within a species while "divergent" was used between species.

protein molecules at one time but also for maintaining genetic variability within a chromosome.

The distribution of divergence between two closely related species in the 5′ region, coding region and 3′ region was quite similar to that of variability within species, although the degree of difference was 8.5 times larger, suggesting that the evolutionary forces lead to divergence in a manner similar to that by which they maintain genetic variability within species at least at the histone gene loci. This is consistent with the explanation of polymorphism by the neutral theory as a phase of molecular evolution (KIMURA and OHTA 1971).

Differences in the silent sites in the coding region was greater than that in the total coding region and virtually the same as that in the 3′ region. Selective constraint in the 3′ region may be very weak or even neutral, since many deletions, insertions and substitutions were found in this region. That variation was greater and the evolutionary rate higher in less constraint sites such as silent sites, than in the amino-acid-altering region suggests the occurrence of purifying selection also in the protein coding region of this multigene family as reported by KREITMAN (1983) at the *Adh* locus. There were greater variability and divergence in the 3′ region or silent sites than 5′ region. There may possibly be regulatory sites sensitive to environmental change and important to the tissue specific or developmental expression of genes. Some differences noted in the 5′ region may affect the fitness of individuals.

The degree of variability within a *D. melanogaster* population and that of divergence between *D. melanogaster* and *D. simulans* was compared between histone H3 and *Adh* loci (Figure 6). The distribution pattern of variation was somewhat different between *Adh* and H3, although the total degree of variation was virtually the same. Variability and divergence in

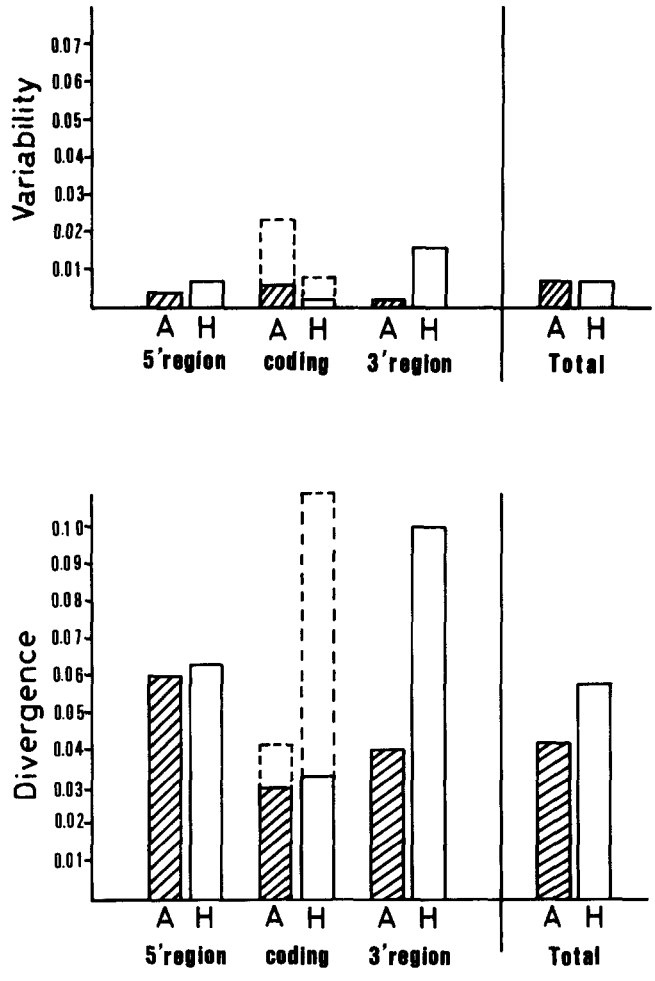

FIGURE 6.—Degree of variation and divergence at Adh (A) and histone 3 (H) loci in the 5′ region, coding region and 3′ region. Dotted boxes indicate the variation and divergence of the silent site.

the 5′ region were quite similar in the two loci. However, the degree of the variability in the coding region of *Adh* appeared more than that in H3, and the degree of the divergence less. This tendency be-

comes more apparent by comparison with the differences in the silent sites. The data of 3′ region are complicated due to another coding frame found in the 3′ region of *Adh* (KREITMAN and AGUADE 1986). There are at least two possible explanations for the above differences in variability and divergence in the coding region. First, the differences may arise from the particular nature of the loci. Different loci may be subject to different selection pressure. But on this basis, it is difficult to explain the larger variability and smaller divergence at the *Adh* locus. Secondly, the differences may result from gene multiplicity. The *Adh* gene is considered a single copy gene, but the histone gene comes from a tandemly duplicated multigene family. If the modes of selection, mutation, unequal crossing over and gene conversion differ in these gene systems, the amount of variability and divergence will of course, differ. It may be reasonable to speculate that homogenizing mechanisms quite likely occur more frequently within a species in a multigene family but that genetic divergence occurs independently, resulting in greater differences between species due to weaker selection pressure of the multigene family. It would thus follow that multigenes are more homogeneous than single copy genes. Other single copy gene and multigene families should be examined for confirmation of this.

From the analysis of the phylogeny of histone genes and a comparison of genetic differences at the three levels, it has been concluded that the histone gene family evolved in a concerted fashion. Several homogenizing mechanisms have been proposed to explain this mode of evolution, such as unequal crossing over (PETES 1980; SZOSTAK and WU 1980), gene conversion (SCHERER and DAVIS 1980) and molecular drive (DOVER 1982). The results of restriction mapping and phylogeny of histone clones in the previous study showed closely linked histone units to be closely related (MATSUO and YAMAZAKI 1989). This situation must surely have come about through a homogenization process. The behavior of mutant genes in a family expected on an unequal crossing-over model (OHTA 1980) is quite consistent with this view. The unequal crossing-over model also suffices to explain variation in the repetition of a simple sequence (SLIGHTOM, BLECHL and SMITHIES 1980).

OHTA (1980, 1982, 1983a, b) and NAGYLAKI (1983a, b, 1984) and NAGYLAKI and PETES (1982) developed a mathematical theory of multigene families. OHTA (1980) first studied a model of unequal crossing over, and formulated the behavior of a mutant in a family by a double diffusion process *i.e.*, unequal crossing over (intrachromosomal) and random genetic drift. The model of gene conversion may be treated more exactly, using the average changes in identity coefficients; these coefficients are the probability of identity of genes at different loci on one chromosomes (C_1), the probability of identity of two genes taken from different loci of two homologous chromosomes of a population (C_2) and the probability of identity of genes at the same locus (f). The important parameters are: effective population size (N), family size (n), conversion rate per gene per generation (λ), recombination rate per generation (β) and mutation rate per generation (v). It has been shown that there is a rough correspondence between the model of unequal crossing over and that of gene conversion. A most important quantity in both models is the rate of steady decay of genetic variability within a single multigene family. This becomes $2\lambda/(n-1)$ in the conversion model, and $m\gamma/n^2$ in the unequal crossing over per gene family, and m is the mean number of genes shifted at unequal crossing over (OHTA 1983a). These two values are roughly interchangeable for predicting gene identity (OHTA 1983a, b).

The equilibrium values of C_1, C_2, and f were presented in OHTA (1982). If we put the values of $N\lambda$, $N\beta < n$ and Nv into the formula (8) of OHTA (1982), the expected values of C_1 and C_2 can be calculated.

TABLE 3

Equilibrium identity coefficients

		Nonallelic identity		
$N\lambda$	$N\beta$	Within chromosome C_1	Between chromosomes C_2	Allelic identity f
15	0.008	0.989	0.985	0.985
	0.08	0.956	0.953	0.954
	0.8	0.818	0.816	0.819
	8	0.711	0.711	0.716
	80	0.693	0.693	0.698
50	0.008	0.997	0.993	0.993
	0.08	0.986	0.982	0.982
	0.8	0.908	0.905	0.906
	8	0.752	0.751	0.753
	80	0.700	0.700	0.702
500	0.008	1.000	0.996	0.996
	0.08	0.999	0.995	0.995
	0.8	0.987	0.984	0.984
	8	0.909	0.907	0.907
	80	0.753	0.752	0.752
5000	0.008	1.000	0.996	0.996
	0.08	1.000	0.996	0.996
	0.8	0.999	0.995	0.995
	8	0.988	0.984	0.984
	80	0.910	0.907	0.907
7000	8	0.991	0.987	0.987
	80	0.929	0.926	0.026
70000	80	0.991	0.987	0.987

The observed values of C_1 and C_2 in the 3′ region were 0.991 and 0.985, respectively, and $n = 110$. If Nv is larger than 0.001, a larger value of $N\gamma$ or a smaller value of $N\beta$ is needed to explain the above data.

The results of some calculations are shown in Table 3. The observed values of C_1 and C_2 in the 3′ region of H3 locus are 0.991 and 0.985, respectively. If the interchromosomal recombination rate is very low ($N\beta \ll 1$), the variation maintained in the histone family can be explained by the unequal crossing-over model (OHTA 1980, 1982, 1983a, b), otherwise, the observed data must be explained on the basis of other mechanisms of homogenization, such as biased gene conversion or duplicative transposition. The values of Nv, $N\lambda$ and $N\beta$ should be measured to facilitate clarification of this point.

The members of a histone multigene family are not completely uniform. About 50% of genetic variation within a population is maintained between loci within a chromosome. Some variation may be homogenized by certain mechanisms. Some members may be the raw material for gene differentiations. To discuss the maintenance mechanism of variation and divergence in a multigene family, measurement must be made of the conversion rate, unequal crossing-over rate, mutation rate and selection coefficient of the locus. Information regarding various loci in various environments must also be available for the study of adaptive evolution in organisms.

We are grateful to T. OHTA for her valuable suggestions, and T. MUKAI for his encouragement. Thanks are also due to K. SAIGO for his many valuable comments, technical suggestions and providing the histone probe, F. TAJIMA for assisting in constructing the phylogeny and M. HATTORI for showing us how to conduct DNA sequencing efficiently. We thank S. YOKOYAMA and M. M. GREEN for editing the manuscript.

LITERATURE CITED

AQUADRO, C. F., S. F. DESSÉ, M. M. BLAND, C. H. LANGLEY and C. C. LAURIE-AHLBERG, 1986 Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. Genetics **114:** 1165–1190.

ARNHEIM, N., 1983 Concerted evolution of multigene families, pp. 38–61 in *Evolution of Genes and Proteins*, edited by M. NEI and K. KOEHN. Sinauer Associates, Sunderland, Mass.

BENTON, W., and R. DAVIS, 1977 Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196:** 180–182.

BODMER, M., and M. ASHBURNER, 1984 Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of Drosophila. Nature **309:** 425–430.

CHILDS, G., C. NOCENTE-MCGRATH, T. LIEBER, C. HOLT and J. A. KNOWLES, 1982 Sea urchin (*Lytechinus pictus*) late-stage histone H3 and H4 genes: characterization and mapping of a clustered but nontandemly linked multigene family. Cell **31:** 383–393.

COEN, E., T. STRACHAN and G. DOVER, 1982 Dynamics of concerted evolution of ribosomal DNA and Histone gene families in the *melanogaster* species subgroup of Drosophila. J. Mol. Biol. **158:** 17–35.

COHN, V. H., M. A. THOMPSON and G. P. MOORE, 1984 Nucleotide sequence comparison of the *Adh* gene in three Drosophilids. J. Mol. Evol. **20:** 31–37.

DOVER, G., 1982 Molecular drive: a cohesive mode of species evolution. Nature **299:** 111–116.

GOLDBERG, M., 1979 Sequence analysis of Drosophila histone genes. Ph.D. thesis, Stanford University, Stanford, Calif.

HATTORI, M., S. HIDAKA and Y. SAKAKI, 1985 Sequence analysis of a KpnI family member near the 3′ end of human β-globin gene. Nucleic Acid Res. **13:** 7813–7827.

HOHN, B., and K. MARRAY, 1977 Packaging recombinant DNA molecules into bacteriophage particles *in vitro*. Proc. Natl. Acad. Sci. USA **74:** 3259–3263.

IKENAGA, H., and K. SAIGO, 1982 Insertion of a movable genetic element, 297, into the T-A-T-A box for the H3 histone gene in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **79:** 4143–4147.

KARN, J., S. BRENNER and L. BARNETT, 1983 New bacteriophage lambda vectors with positive selection for cloned inserts. Methods Enzymol. **101:** 3–19.

KARP, P. W., 1980 The organization of the histone genes in *Drosophila melanogaster*. Ph.D. thesis, Stanford University, Stanford, Calif.

KIMURA, M., and T. OHTA, 1971 Protein polymorphism as a phase of molecular evolution. Nature **229:** 467–469.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–417.

KREITMAN, M., and M. AGUADE, 1986 Excess polymorphism at the Adh locus in *Drosophila melanogaster*. Genetics **114:** 93–110.

LANGLEY, C. H., E. A. MONTGOMERY and W. F. QUATTLEBAUM, 1982 Restriction map variation in the Adh region of Drosophila. Proc. Natl. Acad. Sci. USA **79:** 5631–5635.

LIFTON, R. P., M. L. GOLDBERG, R. W. KARP and D. S. HOGNESS, 1977 The organization of the histone genes in *Drosophila melanogaster*: functional and evolutionary implications. Proc. Cold Spring Harbor Symp. Quant. Biol. **42:** 1047–1051.

MANDEL, M., and A. HIGA, 1970 Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. **53:** 159–162.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MANIATIS, T., R. C. HARDISON, E. LACY, J. LAUER, C. O'CONNELL, D. QUON, G. K. SIM and A. EFSTRATIADIS, 1978 The isolation of structural genes from libraries of eukaryotic DNA. Cell **15:** 687–701.

MATSUO, Y., and T. YAMAZAKI, 1989 *t*-RNA derived insertion element in histone gene repeating unit of *Drosophila melanogaster*. Nucleic Acids Res. **17:** 225–238.

NAGYLAKI, T., 1983a Evolution of a large population under gene conversion. Proc. Natl. Acad. Sci. USA **80:** 5941–5945.

NAGYLAKI, T., 1983b Evolution of a finite population under gene conversion. Proc. Natl. Acad. Sci. USA **80:** 6278–6281.

NAGYLAKI, T., 1984 Evolution of multigene families under interchromosomal gene conversion. Proc. Natl. Acad. Sci. USA **81:** 3796–3800.

NAGYLAKI, T., and T. PETES, 1982 Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics **100:** 315–317.

OHTA, T., 1980 *Evolution and Variation of Multigene Families*. Springer-Verlag, New York.

OHTA, T., 1982 Allelic and nonallelic homology of a supergene family. Proc. Natl. Acad. Sci. USA **79:** 3251–3254.

OHTA, T., 1983a Time until fixation of a mutant belonging to a multigene family. Genet. Res. **41:** 47–55.

OHTA, T., 1983b On the evolution of multigene families. Theor. Popul. Biol. **23:** 216–240.

PARDUE, M. L., L. H. KEDES, E. S. WEINBERG and M. L. BIRNSTIEL, 1977 Localization of sequences coding for histone messenger RNA in the chromosomes of *Drosophila melanogaster*. Chromosoma **63:** 135–151.

PETES, T. D., 1980 Unequal meiotic recombination within tandem

arrays of yeast ribosomal DNA genes. Cell **19:** 765–774.

RIGBY, P. W. J., M. DIECKMANN, C. RHODES and P. BERG, 1977 Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113:** 237–251.

SAIGO, K., L. MILLSTEIN and C. A. THOMAS, JR., 1981 The organization of *Drosophila melanogaster* histone genes. Cold Spring Harbor Symp. Quant. Biol. **45:** 707–710.

SANGER, F., A. R. COULSON, B. G. BARRELL, A. J. H. SMITH and B. A. ROE, 1980 Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143:** 161–178.

SANGER, S., S. NICKLEN and A. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHERER, S., and R. W. DAVIS, 1980 Recombination of dispersed repeated DNA sequences in yeast. Science **209:** 1380–1384.

SLIGHTOM, J. L., A. E. BLECHL and O. SMITHIES, 1980 Human fetal G_γ- and A_γ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21:** 627–638.

SNEATH, P. H. A., and R. R. SOKAL, 1973 *Numerical Taxonomy.* W. H. Freeman, San Francisco.

SOUTHERN, E., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–518.

SZOSTAK, J. W., and R. WU, 1980 Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae.* Nature **284:** 426–430.

VIEIRA, J., and J. MESSING, 1982 The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:** 259–268.

YAMAZAKI, T., S. KUSAKABE, H. TACHIDA, M. ICHINOSE, H. YOSHIMARU, Y. MATSUO and T. MUKAI, 1983 Reexamination of diversifying selection of polymorphic allozyme genes by using population cages in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **80:** 5789–5792.

YAMAZAKI, T., Y. MATSUO, Y. INOUE and Y. MATSUO, 1984 Genetic analysis of natural populations of *Drosophila melanogaster* in Japan. I. The amount of protein polymorphisms, lethal genes, sterility genes and inversion polymorphisms, and linkage disequilibrium. Jpn. J. Genet. **59:** 33–49.

Communicating editor: D. CHARLESWORTH